DOCUMENT RESUME

ED 322 158 TM 015 248

TITLE	Educational Quality Indicators: Taking Stock. Proceedings of the Conference (Los Angeles, California, October 12-13, 1989).
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.; Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Dec 89
NOTE	22p.
PUB TYPE	Collected Works - Conference Proceedings (021) -- Reports - Evaluative/Feasibility (142)
JOURNAL CIT	CRESST Evaluation Comment; Dec 1989
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Abstracts; *Accountability; Achievement Tests; Conference Papers; *Educational Assessment; *Educational Quality; Elementary Secondary Education; High Risk Students; International Studies; Local Issues; National Programs; Performance; School Districts; School Restructuring; State Programs
IDENTIFIERS	*Educational Indicators; *Quality Indicators

ABSTRACT

An overview of an international conference held on the campus of the University of California at Los Angeles (UCLA) to take stock of the development and use of educational quality indicator systems at the local, state, national, and international levels is provided. Major implications and findings of the education summit held at the University of Virginia (Charlottesville), September 27-28, 1989, by President Bush and the nation's governors are discussed in the opening address by Emerson J. Elliott entitled "Accountability in the Post-Charlottesville Era." Educational reform in the area of accountability; indicators used by the National Center on Education Statistics; setting national goals and assigning accountability; measuring processes; and issues related to accountability and indicator systems, i.e., controls, and content, are examined; and an agenda for the future is briefly considered. Following this, summaries of six plenary sessions are provided. Because all groups in the education community must be involved in the dialog about the status and improvement of indicator systems, each panel of speakers included policy makers, practitioners, and researchers. The session topics include: (1) National and State Issues: The Role of NAEP; (2) State and District Issues: The Role of Indicators and Assessment in School Reform and School Restructuring; (3) National and State Issues: The Impact of Commercial Achievement Tests; (4) Performance Assessment: Implications for Large-Scale Assessment and Indicators; (5) Accountability and At-Risk Students; and (6) International Educational Indicators: Their International and National Roles. In conclusion, conference participants' responses to a questionnaire that asked them to specify the benefits and dangers of indicator systems and to identify implications for policy, practice, and research are identify implications for policy, practice, and research are summarized. Sixty-one publications available from the UCLA Center for the Study of Evaluation are listed. (RLC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. C. BEER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED322158

Educational Quality Indicators: Taking Stock. Proceedings of the Conference (Los Angeles, California, October 12-13, 1989)

TM015248

BEST COPY AVAILABLE

EVALUATION COMMENT

UCLA CSE

In collaboration with the University of Colorado;
NORC, University of Chicago; and Arizona State University

In This Issue ...

The public's call for credible information about educational quality and the emphasis President Bush and the governors of the fifty states have placed on information and assessment issues point to the increasingly prominent role that educational indicators will play in mobilizing educational reform. The risk in not responding to such concerns is high: If the education community fails to advance a responsible national agenda for quality indicators, America's standing in the world economy could be impaired and national pride could suffer.

This concern impelled CRESST to sponsor an international conference, "Educational Quality Indicators: Taking Stock." The theme of the conference emphasized the notion of "taking stock" of the development and use of indicator systems at the local, state, national, and international levels. The conference brought together policy makers, practitioners, and researchers to address these issues in a forum designed to encourage an exchange of ideas among the participants. This volume of *Evaluation Comment* presents the proceedings of the conference, which was held on the UCLA campus on October 12 and 13, 1989.

Emerson Elliott, Director of the National Center for Education Statistics, opened the conference; his address introduced many of the issues that participants discussed during the two days of plenary and working group sessions. Dr. Elliott's address is presented below; highlights of the plenary sessions follow.

Conference participants also responded to a questionnaire that asked them to specify the benefits and dangers of indicator systems and to identify implications for policy, practice, and research. A summary of their responses follows the plenary session synopses.

Leigh Burstein
Eva L. Baker
CRESST

ACCOUNTABILITY IN THE POST-CHARLOTTESVILLE ERA

Emerson J. Elliott
Acting Commissioner
National Center for Education Statistics

INTRODUCTION

On September 27 and 28 President Bush convened an historic education summit with the Nation's Governors at the University of Virginia in Charlottesville. This was only the third time in our 200-year history that a president had called such a summit with the governors and the first time such a conference was a call to action in the education arena.

In his concluding remarks President Bush stated, "...we unanimously agree that there is a need for the first time in this nation's history to have specific results-oriented goals. We recognize the need for...accountability for outcome-related results." This serves as the backdrop for my remarks to you today and, indeed, for the whole of this CRESST conference on education quality indicators.

Again, Not Anew

Last May, our national scribe for state government activities in education, Chris Pipho, was already telling us that education accountability was coming not anew, but again (*Phi Delta Kappan*, May 1989). Chris said that it was here before, in the 1970s, when some 31 states had enacted legislation dealing with accountability and more than one-third of those states were using a systems approaches: "The specific topics covered by this early legislation included: assessment of student achievement, evaluation of programs, setting goals for education, specifying objectives for learners, PPBS (planning, programming, budgeting system), MBO (management by objectives), MIS (management information systems), uniform accounting systems, and performance accreditation systems."

Does this sound like 1989? There was a difference, Chris reminds us. Accountability in the 1970s was aimed at "more bang for the buck...the application of the tools of business management to education...[promising]...a new era of efficiency."

Focus on Achievement

Moving on to the present, Chris states, "Most of the legislation enacted during these early years is still on the books, but state

policy makers seem to be moving toward a new brand of accountability that is more closely tied to instruction. Measuring student performance (or the lack of it) and assigning responsibility for improving the situation seem to best describe the goals of the new model."

This focus on achievement was also reflected in a recent Gallop Poll in which 70% of Americans questioned favored requiring schools "to conform to national standards and goals." And, Ernest L. Boyer has observed: "I think we've gone about as far as we can go in the current reform movement dealing with procedural issues." By establishing national academic standards and exams, schools "would be held accountable for outcomes rather than the current situation of heavy state regulation that 'nibbles' them to death over procedures."

Thus, the focus has moved from a system where an individual, or institution, is accountable for procedural or process standards to one demanded by parents and other taxpayers where accountability is expressed in terms of student achievement and outcomes.

A NEW WAVE OF REFORM

In the last few months, we have seen reports in the pages of our newspapers and trade journals about what one state and then another is doing to "reform" education and to make it more "accountable":

- Thirty state legislatures scheduled discussions of educational accountability for their 1989 sessions (Chris Pipho, *Phi Delta Kappan*).
- New Jersey is monitoring school districts for performance in several areas, including such outcomes as student performance in basic skills, attendance, and school-community relations. The state also plans to issue an annual "School Report Card" to publicize progress and achievements in each individual school. Both actions are intended to reward good performance and expose poor performance. For failing districts, the state is using its authority to declare educational bankruptcy and take over in cases of particularly severe problems (i.e., Jersey City).
- Governor's commissions in Kansas and Maryland have proposed accreditation systems that would measure student outcomes rather than rely solely on measures of educational input.
- In Iowa, the chief state school officer and the governor are proposing exit tests for all high school seniors that would measure knowledge and skills, including problem-solving skills, in mathematics, science, and writing.
- California is developing a report card that will track indicators of school-level performance in a dozen areas, including achievement, dropout rates, teaching loads, systems for teacher evaluation and training, quality of textbooks, and several others.

These are merely examples. A healthy majority of states can cite recent policy initiatives that, in one way or another, try to hold schools accountable for the quality of instructional processes and outcomes.

Changes in the Reform Movement

One remarkable thing about these reforms is the continuing momentum that comes from state political leadership. The reforms of the 1980s—unlike ones that were put into place during the 1970s—have engaged governors and legislators to an unprecedented degree. Indeed, many political futures have been, or are, on the line for improvements in education. There has been a ceaseless agitation and action in the states at least since the 1983 *A Nation at Risk* report. There are few topics on the political agenda that have shown such lasting power. This is a measure of the importance the U.S. public attaches to education as the role of education in our economy becomes increasingly evident.

A second remarkable aspect of these school reforms is a shift in approach since 1983. The first wave of reform emphasized many elements that educators had talked about for years—higher teacher pay, merit pay, tougher certification requirements, tougher graduation requirements. We find that teacher pay, in constant dollars, had passed its 1973 peak by the 1987-88 school year. Per pupil expenditures nearly quadrupled in constant dollars between 1949-50 and 1986-87. And, states are paying a bigger share of the bill, now about 50% annually. Yet, an ever impatient public, not about to wait a generation for better results, points to stubbornly disappointing achievement.

The more recent interventions, such as the examples I cited previously, are varied (showing that Federalism is very much alive) but place greater emphasis—as did the Charlottesville conference—on what political leaders want students to accomplish.

A third element of the new wave of reforms—dare we say it—is that education researchers have had more influence: current reforms place more emphasis on content of the curriculum, learning exposure time, higher order thinking skills, the site of decision making (at the school level), and the role of teachers.

And, a fourth element, one I cannot pass up, is that the reform movement has sparked an interest in comparable data that has turned into a conflagration at the National Center for Education Statistics!

"REFORM" AND NCES

With the Council of Chief State School Officers (CCSSO), the National Center for Education Statistics (NCES) has studied common terms and definitions and is working to standardize them. We have a new state/federal cooperative statistics program intended to help make state data more comparable and uniform. Congress has asked us to convene an advisory panel to make recommendations for education indicators. In preparing these remarks, I have drawn on the background papers written for that panel. (In this regard, I'd especially like to acknowledge the work of Brenda Turnbull, from Policy Studies Associates in Washington, and John Ralph of the NCES staff.)

The Congress has also requested new annual national data collections and an annual report on school dropouts. In addition, they have added state-representative reports, on a pilot basis, to the National Assessment of Educational Progress.

NCES appropriations are now three times what they were in fiscal 1986, so that new data collections are possible in such areas as schools and staffing, the eighth-grade longitudinal cohort, state assessments under NAEP, college faculty, and student financial aid. Our activities in the international arena have expanded as well, with international assessments in science and mathematics, literacy, and an OECD indicators project.

Indicators Used by NCES

NCES's most direct involvement in accountability and quality indicators has been in recent editions of its annual publication, the *Condition of Education*. I'm proud of this work and pleased to note

many of you have advised us on the contents of that volume at one time or another.

This year's report displays thirty simple measures and data relationships at the elementary and secondary level showing changes over time; comparing or contrasting subpopulations, regions, or states; or describing characteristics of students from different backgrounds. We assert, I think with reason, that these indicators are the most valid and representative education statistics available in America today for the subjects and issues they portray.

However, I always feel that some statement is required to explain NCES's professional role in making the selection of data to be displayed there. This year's report includes a statement that indicators "represent a consensus of professional judgment on the most significant national measures of the condition and progress of education at this time, but tempered, necessarily, by the availability of current and valid information."

We have many debates at NCES about what kinds of data analysis are appropriate for a statistical agency to do. Where do analysis of data and relationships within the data spill over into policy advocacy or bias? Strangely, we don't have debates of that sort about the indicators we have selected to display, and yet each one implies that the relationship being described has an important value in education. For example:

- There are seven measures of student performance in reading, mathematics, science, history and literature, and computer competence.
- There is an indicator on the proportion of high school graduates who have taken the "new basics" courses advocated in *A Nation at Risk*.
- There is a measure of unemployment rates that compares 20- to 24-year-olds who have graduated from high school compared with those who have dropped out.
- Expenditures per pupil over time are displayed.
- There is an index of financial "effort" that relates per pupil expenditures to per capita wealth.

I'll agree that there is a difference between values and specific goals. Still, I have thought these indicators activities were probing the limits of appropriate NCES activity and I have eagerly anticipated the work of the indicators panel to sanction them or devise different ones.

THE PRESIDENT AND THE GOVERNORS

All of that was before Charlottesville. It seems like a different era now, even though the summit was just "yesterday."

For more than the three decades I have been in Washington, the government's role in education has been bounded by statutory language meant to limit federal activity that would "control" curriculum. In fact, the U.S. Department of Education organization act includes language that prohibits offices of the department from exercising:

> any direction, supervision, or control over the curriculum, program of instruction, administration, or personnel of any educational institution, school, or school system, over any accrediting agency or association, or over the selection or content of library resources, textbooks, or other instructional materials by any educational institution or school system, except to the extent authorized by law.

And, while no one has rescinded this prohibition, nor is anyone about to, it reflects a different sort of Federalism from the one that appeared to prevail in Charlottesville.

Setting National Goals and Assigning Accountability

In the Jeffersonian Compact issued by President Bush and the governors, a new context is described. It includes these statements:

> Education has always been important, but never this important because the stakes have changed: Our competitors for opportunity are also working to educate their people. As they continue to improve, they make the future a moving target.

The Compact goes on to make an assertion no one could have predicted a few months—even a few weeks—ago:

> We believe that the time has come, for the first time in U.S. history, to establish clear, national performance goals, goals that will make us internationally competitive.

The goals themselves are to be formulated through joint action of the National Governors Association and the U.S. Federal Government and are to engage teachers, parents, administrators, school board members, elected officials, business, labor, and the general public. They are to deal with:

- Readiness of children to start school.
- Performance on international achievement tests, especially math and science.
- Improvement in academic performance.
- Reduced dropout rates.
- Functional literacy of adults.
- Training level for a "competitive" work force.
- Supply of qualified teachers.
- Supply of up-to-date technology.
- Safe, disciplined, drug-free schools.

The Compact describes a "Federal-State Partnership" and the U.S. Federal Government's role, including—not surprisingly—the provision of: "...good information on the real performance of students, schools and states..."

It concludes with another ground-breaking assertion:

> As elected chief executives, we expect to be held accountable for progress in meeting the new national goals and we expect to hold others accountable as well. When goals are set and strategies for achieving them are adopted, we must establish clear measures of performance and then issue annual Report Cards on the progress of students, schools, the states, and the Federal Government.

Measuring Progress

In his own remarks, President Bush said, "To get results, we will need a new...report card...we need to know just how much progress we're making. We've always measured our progress against our past performance. We must now evaluate ourselves on a tougher grading curve—one that includes the other major industrial nations." What a challenge for the participants in this conference!

The calls for "better" data—more descriptive, more quickly produced, more detailed—will be loud and insistent. The visibility of reporting systems will be heightened. The press will eagerly await each new report and will spread columns of words and charts throughout our newspapers.

There will be countless spots on the evening TV news and numerous authorities will be interviewed on PBS and network programs. There will be still more panels, commissions, probably even more money—maybe CRESST will even have more to do.

Evaluation Comment — Page 3

Certainly, its advice will be sought. And so, whatever Eva's reasons were for scheduling this conference, "quality indicators"—as we have found with NAEP's taking on a state dimension—are now transformed into a "high stakes" national issue.

The numbers will matter. They will matter, of course, to the governors and to President Bush, who have set themselves on such a bold path and will be judged by their accomplishments. They will matter to educators who will be expected to find successful ways to reach new performance goals. They will matter to students, parents, and taxpayers who have vital stakes in what American education accomplishes.

They will matter to our panel on education indicators who well may question what their role can be in such a changed environment. But, if the numbers matter so much, then the shortcomings inherent in the development and implementation of indicator systems for accountability matter all the more.

These shortcomings were as visible as Central Park when you fly over Manhattan—it is there clearly enough, but it is a little remote. Now they have the added attributes of a walk through Times Square—immediate, bright, brassy, full of light and life, magnetic, sometimes frightening, and even dangerous. Therefore, those challenges must now be a direct subject of this conference and an abiding concern of those of us who work with education information. It is our role to be introspective about appropriate uses of indicators and about the pitfalls of misapplications. It is our duty to ensure in every way we can that data are clearly explained and that the public and policy maker are made aware of inappropriate uses.

ISSUES RELATED TO ACCOUNTABILITY AND INDICATOR SYSTEMS

I would like to describe four issues that are especially problematic for accountability systems in a "high stakes" environment. They are: corruptibility, consequences, controls, and content.

Corruptibility

Corruptibility has two facets. One facet is measurement that perverts the very process it is supposed to report on and improve. There are often unintended consequences of measurements that have sanctions attached to them. For example, if we place a priority on reducing the grade-level retention of students, we will undoubtedly see fewer failing grades and, in all likelihood, more social promotion. The effects on students could be negative, if there are any effects at all. Sometimes multiple measures can help avoid this particular problem.

A second facet of corruptibility is measurement that is deliberately misleading. The work of Dr. John Cannell has effectively advanced the claim that some schools cheat on tests. Officials at the state level have known for years that schools sometimes either selectively suppress the lowest individual scores or, even worse, allow answers to fall into students' hands.

The lesson here is that researchers and policy makers need to direct more attention to maintaining the integrity of testing and assessment procedures. This matter is of major and continuing attention to NAEP, where we have acted to acquire valid and comparable data through:

- Consensus process to develop common procedures.
- Specific quality control requirements in sampling, security, and other areas.
- Monitoring and evaluation of actual administration of the assessments.

Consequences

A second issue has to do with the consequences embedded in accountability systems. Too often, I fear, political leaders have not been aware of the consequences of particular reforms or have glossed over conflicting purposes. This creates a potential for indicators to report results quite different from the proponents' intent.

The current debate in Ohio over a four-tiered system of diplomas nicely illustrates this problem. Beginning with the class of 1994, the system would use tests to determine whether each student would receive a "certificate of attendance" or one of three types of diplomas, ranging upward to a "diploma of distinction." An advocate of the proposed system, State Senator Eugene Watts, says, "What we're doing is driving curriculum. We are demanding accountability." (That is—holding schools accountable.)

His chief opponent, Representative Ron Gerberry, counters with an argument at the individual level: "I don't think this is the proper time to stigmatize students. Students who are average or a little bit better than average will never graduate with distinction or commendation, but that doesn't make them less able to achieve a bachelor's degree and be successful."

The opposition is prevailing in the House of Representatives, which has voted, 97 to 1, to repeal the law.

This is an example—at least as viewed from Washington—of legislative failure to be clear as to who is being held accountable for what. If the student is being held responsible for a school failure, then there is a problem with the accountability system. Still, for the student not to be informed about his or her level of achievement is a consequence too.

Somewhat surprisingly, the lively debates over the initial design of accountability systems do not necessarily foreshadow equally lively debates over the actual consequences.

In a CRESST study of state and local programs of competency testing, Mary Catherine Ellwein, Gene Glass, and Mary Lee Smith found that officials focused little or no attention on the ultimate passing rates on these tests. While initial failing rates did attract policy attention, the eventual, cumulative passing rate was seldom made public. When it was, the purpose was simply to hail it as evidence of the test's benefits. Moreover, these researchers found that the research and evaluation community has done little to examine the longer-run effects of testing programs on individuals or school systems.

Controls

A third issue is controlling for student background in reporting results. We call this "fair comparisons" in NAEP. When or whether to "adjust" is the issue here and it has been a perplexing one for the National Assessment Governing Board.

One dimension of this question concerns instrument sensitivity. Rather than simply reporting on which schools are meeting uniform goals and which are not, we must measure progress towards goals. This requires taking into account where the school began or what its educational needs are.

Drawing on the experience of the states that already adjust their measurement or reporting, we need to gather information on the effects of different ways of "comparing likes to likes." South Carolina, for example, compares school performance with an expected score calculated on the basis of each school's past performance and the current performance of other, similar schools.

The student level, though, is a second dimension of comparison. Although it is appropriate to hold schools to flexible stan-

dards of progress, the resources available to students and the expectations for their performance should all be geared towards "equal intellectual opportunity."

This issue is now becoming familiar to policy makers, but there is a high risk of misunderstanding by the public. Researchers and policy makers need to work together to clarify the different purposes of school and student measures. This is another case where multiple reporting—different types of measures that extend the notion of "fair comparisons"—is to be preferred over simplistic averages.

I am convinced that policy makers and the press can understand such things.

Content

A final issue is the content of indicators—what skills are tested in a performance indicator system and what descriptive process measures are included along with outcome measures.

Students need, and schools teach, many different skills. If indicator systems are to help focus energies of the educational system on important skills, then important skills must be measured. Increasingly, critics of multiple-choice tests are persuading policy makers that these tests unduly narrow the focus of measurement to a limited range of "basic skills"and can have adverse consequences for the curriculum as well. Accordingly, using such tests may inadvertently reduce students' opportunities to learn the very skills they will need most in tomorrow's society, such as breaking a complex problem down into its components.

Although our technology for conventional assessment is quite sophisticated at this point, several different approaches that are now under development offer models worth consideration for wider application in accountability systems. CRESST is, of course, a major national resource on this matter. Some examples of approaches now being developed are:

- The Vermont program will include the submission of student portfolios.
- New York has a statewide test in fourth-grade science that asks students to conduct a short experiment and report the results.
- The Connecticut Assessment of Educational Progress has just begun to develop an assessment of performance in math and science that uses a series of tasks that may take students, individually and in groups, as long as a semester to complete.
- The Department of Education in England and Wales is developing a whole battery of materials for performance assessment across the curriculum.

There are many other examples of new developments, of course. However, "under development" does not equal "proved reliable, objective, and ready for use in large-scale assessment."

Another dimension of content concerns how measures of school process should be used in an indicator system. We might use such measures—measures of teacher effectiveness, pupils' opportunity to learn, school climate or orderliness—as leading indicators, just as economists try to forecast the gross national product by looking at selected input, or leading indicators, periodically.

AGENDA FOR THE FUTURE

These issues give us a solid agenda of tasks for researchers and policy makers in this post-Charlottesville, Jeffersonian Compact era. The issues I've described—corruptibility, consequences, controls, and content—must be central to the research community. Policy makers can't work these out alone. Such confusion as the Ohio example illustrates is not good public policy. Together, we need to:

- Reduce the corruptibility of indicators by attending to their unintended consequences and to the mechanics of measurement.
- Address the dilemma of appropriate consequences by thinking through the appropriate rewards and sanctions for both individuals and bureaucracies. Be clear as to who is accountable for what.
- In a similar vein, strive for a balance between rewarding progress fairly and holding all students and schools to high standards.
- Continue to broaden and refine our measures of performance and process.

We also need to fill gaps in the stock of available indicators. For example, we need to direct attention to school readiness and student transitions (home to school and school to work and/or college).

Ironically, the same educational system that easily and routinely grades student performance shows very mixed feelings about grading itself. But President Bush and governors have set us on that course. When we address the challenges responsibly, accountability can only benefit the system, the students, and the nation.

President Bush and the governors concluded their compact with the following: "The time for rhetoric is past; the time for performance is now." The challenge is to us.

EDUCATIONAL QUALITY INDICATORS: TAKING STOCK

Summaries of Plenary Sessions

Each of the six plenary sessions was designed to focus on an important issue related to the development and use of educational quality indicators. Session topics were chosen to reflect the fact that all levels of the education system—local, state, national, and international—are involved with these issues. Because all groups in the education community must be involved in the dialogue about the status and improvement of indicator systems, each panel of speakers included policy makers, practitioners, and researchers.

In the session summaries that follow, an introduction to the session topic precedes the highlights of speakers' remarks.

NATIONAL AND STATE ISSUES: THE ROLE OF NAEP

NAEP has been given a new charge: to examine the viability of providing achievement data that can be compared at the state level. The National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES), who are responsible for overseeing the implementation and conduct of

NAEP, have placed considerable emphasis on ensuring that the state trials planned for 1990 and 1992 will afford full opportunities to determine the feasibility and viability of obtaining comparable data. The operational contract for state NAEP has been awarded to the Educational Testing Service; the evaluation of state NAEP has been awarded to the National Academy of Education.

Various organizations including the Council of Chief State School Officers (CCSSO) have supported efforts to produce and report such state-level data, and public and corporate support has been expressed as well. Pockets of policy analysts and educational practitioners also have indicated that they would back this effort.

However, resistance to state-level NAEP has been voiced by many members of the education community. Their concerns relate to the technical soundness of the proposed assessment, the burden of additional time and money that the assessment would require, and the ultimate usefulness of the information collected.

Panelists were asked to comment on these issues.

Gordon Ambach, Council of Chief State School Officers

As recently as five years ago, a conference like the CRESST conference on educational indicators would have focused primarily upon minimum competency testing. The changes that have occurred since then have arisen as a result of state and local efforts and may be characterized by an emphasis on accountability linked with a need for increased resources.

- NAEP data is becoming an increasingly high-stakes issue for executives, but because it does not have direct consequences for the students who take NAEP tests, it is a low stakes issue for them. If motivation is low, validity follows suit.
- Integration of national, state, and local assessment systems is imperative. Multiple levels of assessment may overheat the system.
- Because cross-national comparisons may ultimately prove even more important to NAEP's work than state comparisons, the design of appropriate instrumentation must be undertaken.

Chester E. Finn, Jr., National Assessment Governing Board

The essential responsibilities of the NAGB are to decide what will be assessed by NAEP and to establish goals for each subject and grade level. NAEP's approach to assessment is characterized by several important limitations as well as strengths. Among NAEP's limitations:

- Testing is limited to cognitive outcomes and, with only a few exceptions, is restricted to a multiple-choice format.
- It omits foreign language, art, and music.
- It provides no information on important subsets of the population tested, such as the handicapped, the gifted, or students in private schools.
- It is costly in terms of dollars and time.
- NAEP is forbidden by law to make its test items available for use directly by requesting states.

Among its advantages:

- It is increasingly anchored to scales that can remain fixed over time.
- It provides information that the SAT does not, and it tests everyone.
- It yields data useful for formative evaluation and diagnostic purposes as well as summative evaluation.
- It presents its findings in a straightforward manner that is easily understood.

Judith Billings, Washington State Department of Education

Washington is one of many states that opted not to participate in NAEP's early state-by-state trials. The state has been supportive of meaningful assessment efforts for many years and has participated in several large-scale assessments, including regular NAEP testing. The decision to not participate was based partly on the feeling that some basic policy issues remain unresolved. Among them are:

- What purpose will be served by the substantial funding required? What information will be provided that we don't already have?
- Will the states that are involved be denied access to raw data? If so, why?
- Has the meaning of state-by-state comparisons been thought through? Meaningful comparison must take into account such variables as state differences in funding, legislation, student demographics, and teacher certification.
- Are we moving toward a national curriculum? Doesn't this contradict the movement toward local control, decentralization, and site-based decision making?

Dan Koretz, The RAND Corporation

NAEP, despite its extraordinary importance in American education, is being asked to carry too much freight, to serve too many masters. Asking NAEP to serve an accountability role has negative effects:

- Funds channeled to state-by-state NAEP might be used instead to develop assessments with broader scope and improved sampling.
- Teachers may begin to teach to the test, which leads to the degradation of instruction. At the extreme, some schools may not teach the subjects not tested.
- Test items must be held secret to preserve validity, but, as a result, researchers cannot use the results effectively.
- Assessment systems can be designed to be used for information purposes or as a bullwhip—but not both.

STATE AND DISTRICT ISSUES: THE ROLE OF INDICATORS AND ASSESSMENT IN SCHOOL REFORM AND SCHOOL RESTRUCTURING

Many states have launched accountability systems to monitor the effects of their educational reform policies. These systems appear as new student and teacher assessment programs, school report cards, and reports of educational quality indicators.

In some states these systems report findings for the state as a whole; in other states, district- and school-level monitoring and reporting is the standard practice. Some systems incorporate specific rewards and sanctions for districts, schools, teachers, and students. In other systems, the only consequences are those that result from the publication of indicator and assessment information.

Panelists were asked to comment on the viability of using indicator systems in school reform and restructuring.

Lorraine McDonnell, The RAND Corporation

The role of indicators in school reform and restructuring is understood best in the context of a "horse trade" between states, who agree to regulate less, and districts, who promise better performance. Implicit in this arrangement is that no restructuring is possible without accountability. This linkage presents a sound basis for reform, but presents obstacles:

- Although accountability systems are becoming more complex and refined, existing systems are still too dependent on test achievement data.
- Indicators are employed not merely for descriptive purposes but also to reward and punish schools through the allocation of resources and technical assistance, and thus are powerful instruments for changing the behavior of principals and teachers.
- Each proposal for restructuring embodies its own accountability strategy, a situation that presents technical and political barriers. For instance, the decentralization associated with school-based management conflicts with the tendency to centralize curriculum to accord with large-scale assessments.

Roy Truby, National Assessment Governing Board

The role of NAGB is not only to provide descriptive information regarding school achievement, but also to prescribe standards for achievement.

- The prescription of standards requires that achievement and prescription be placed in a common context.
- The actual standards that are set may be less important than is the commitment to achieve those goals.
- Polls indicate that the American public is ready for a national curriculum; however, although this may seem in theory to be a good idea, in practice it may not be desirable or feasible.

Virginia Rosen, Dade County Public Schools

Multilevel indicator systems developed by the Dade County Public Schools collect information that is useful to many people, including educators at the school level:

- Indicators assess some 12 to 15 aspects of school performance.
- Indicators are standardized over time—not based on definitions that may shift from year to year. Variables such as dropout rates, participation in upper-level classes, and performance on Advanced Placement exams are monitored.
- State and national assessment programs may be based on indicators that are important for a political reason alone. If indicators are going to be useful, they must be able to be used effectively at the school site.

Sharon Robinson, National Education Association

The National Education Association works with groups of teachers to identify positive instructional changes that they can implement in their classrooms. The program addresses issues of process and content.

- Teachers should share a common educational vision with their colleagues rather than working in isolation.
- Successful change is evolutionary and is engendered through planning and assessment.
- Teachers must have access to information in order to make sound decisions; this issue is one of the most stubborn and important issues facing proponents of school restructuring.
- Quality indicators must measure input into the educational system, input that includes time, money, staff development, and teachers' access to support. In this regard, the rhetoric of quality indicators far outpaces the reality.

NATIONAL AND STATE ISSUES: THE IMPACT OF COMMERCIAL ACHIEVEMENT TESTS

States have expanded their use of commercial achievement tests in recent years as part of their state testing programs, and districts continue to rely heavily on such tests for program evaluation and monitoring and reporting of test results.

Criticism of the use of commercial tests in these testing programs has been considerable, however, focusing primarily on misleading interpretations of test results, such as those that report that all students are above the national average, as noted in 1987 by John J. Cannell.

Panelists were asked to consider the present role of commercial achievement tests and the impact that these tests have on assessment efforts, and to suggest changes that would maintain or improve their utility in indicator systems.

Robert Linn, CRESST, University of Colorado

A 1989 CRESST survey of test reports from a number of states and school districts showed that more than 50% of students were described as being above the national median. However, several cautions should be noted in conjunction with this finding:

Evaluation Comment — Page 7

- Norm-referenced tests are administered up to 6 or 7 years past their reference year, and successive re-normings have raised the level of difficulty that the norm represents. As a result, older tests show student performance that is generally above the norm.

- Adjustment for the test's reference year accounts for much of the improvement reported in student achievement. Focus should be shifted from norms to the actual performance levels that test scores represent.

- The pressures of accountability have implications for test administration. The context of administration must be understood in full, especially who is tested and who is not.

John Keene, National Computer Systems, Inc.

The proper use of norm-referenced tests must incorporate an understanding of the structure and processing of the tests:

- Several key "control points" have a bearing on the proper use of tests: (a) the method used to report results, in terms of the descriptive measures used and standards of comparison; (b) the selection of test content; (c) aspects of test administration, including the selection of the reference year; (d) sampling procedures used in constructing the norm; and (e) proper interpretation of results.

- Norm-referenced tests have "liquid standards"; that is, such tests are sensitive to instruction and thus are affected easily by instruction that is directed toward test content.

- Test results can be and often are misinterpreted easily.

Floraline Stevens, Los Angeles Unified School District

Although there are problems involved in interpreting the results of commercial norm-referenced achievement tests, they can serve a positive function in a school district such as Los Angeles Unified if they are used properly:

- Contextual factors, such as adverse social factors and high proportions of inexperienced teachers, must be taken into account when test results are discussed.

- Test results can be used to revise curriculum, improve instruction, and determine resource allocation.

- The results from norm-referenced tests can provide important indicators of individual student progress, staff allocation priorities, and students' opportunities to learn important subject matter within individual classrooms.

- Teacher training can encourage teachers to take responsibility for test results and to develop strategies for improving student achievement.

Stanley Bernknopf, Georgia Department of Education

Commercially produced norm-referenced tests have made an impact on Georgia schools in several areas.

- A significant amount of money is allocated to programs on the basis of test results. Tests exert an influence on the curriculum that rivals the influence of state-mandated curricular objectives.

- Achievement outcomes have become an important campaign component for state and local offices, and tests have become a cornerstone of legislation aided at increasing accountability. The pressures arising from reform and accountability policies threaten the state's assessment system's utility for diagnosis and improvement of instruction at the school and classroom levels.

- A two-track system of evaluation may develop, with one set of tests used for accountability purposes and the other used by school personnel.

PERFORMANCE ASSESSMENT: IMPLICATIONS FOR LARGE-SCALE ASSESSMENT AND INDICATORS

Recent proposals for the inclusion of performance assessment activities in testing programs represent a major new force in educational assessment. The development of performance assessment has been spurred by renewed interest in higher order reasoning and problem-solving skills; many educators believe that performance assessment can better reflect these skills—what they think students ought to be learning—than can multiple-choice, pencil-and-paper tests.

Several states, including California, Connecticut, and Michigan, are taking the lead in developing methods for incorporating performance assessment as central features of their testing programs.

Panelists were asked to recount their experiences with alternative assessment approaches and to speculate on their possible integration into large-scale assessment programs and their potential use as educational quality indicators.

Richard Shavelson, University of California, Santa Barbara

Researchers at UC Santa Barbara are developing testing methods for math and science that focus on performance assessment; if successful, the method will enable better understanding of instructional methodology in these content areas.

- For science, the testing method is based on the analysis of three laboratory experiments designed for elementary students. Scoring of the experiments was based on an expert-novice benchmarking procedure. This testing method incorporates computer simulation, modified multiple-choice items, and conventional multiple-choice tests (CTBS).

- For math, a test was developed that asks students to generate mathematical problems with a given set of criteria.

Eva Baker, CRESST, University of California, Los Angeles

Researchers at CRESST are developing a method of assessing higher order thinking in history through performance measures.

- Because thinking about a subject such as history requires active construction, elaboration, and integration of prior knowledge, a performance measure was judged to be the best vehicle for assessment.

- The measure includes a test of prior knowledge, text from primary historical source material, a reading comprehension test, an essay question, a measure of student anxiety, and debriefing questions.

- During assessment, students are asked to read the text of a speech and then write an essay, incorporating their prior knowledge about the topic. Essay raters consider the organization of the essay (whether it is organized by a premise or problem), use of prior knowledge, use of text-based knowledge, use of interrelationships, absence of misconceptions, and overall quality of content.

- In field tests, interrater reliability was found to be extremely high. Validity was determined by comparing students' scores to teachers' expectations of students' performance and scores on standardized instruments.

Judith Torney-Purta, University of Maryland

Researchers at the University of Maryland are using performance assessment to measure teenagers' understanding of foreign policy. Performance tests are being used to elucidate the cognitive structures of student knowledge.

- If tests assess cognitive structures, or maps, and teachers gear their instruction toward maximizing student performance on these tests, the entire teaching-learning process will benefit. Multiple-choice tests are limited because, in general, they do not tap students' cognitive structure.

- A computer simulation is used to assess teenagers' understanding of foreign policy. Through role play, students attempt to solve dilemmas posed in the test. Pre- and posttest interviews are administered to ascertain qualitative differences in student thinking after the role play.

- In pilot tests, students with complex cognitive structures were those who could: (a) suggest multiple solutions; (b) see constraints on proposed solutions; (c) see the implications of action and how they could affect development of alternatives; and (d) rank countries along a complex set of dimensions, including economic status, type of economy, and presence of natural resources.

Ed Roeber, Michigan Department of Education

Performance assessment should be a vital part of large-scale assessment programs and indicator systems, whether at the national, state, or local level. Performance measures can give an entire assessment an added aura of content validity.

- Performance assessment is not a new idea. In 1971, teachers in areas such as physical education, health education, and music were at the fore in calling for performance testing; the teachers of public speaking, writing, and communication followed this lead. However, after initial forays were made with small samples of students, state officials realized that testing every student in this manner was costly.

- Performance assessment can be used to determine whether students can in fact demonstrate the critical skills that we want them to learn.

- Today, data on student achievement is used not only to assess students, but to evaluate education systems and implemented program changes. Because so much weight is given to test results, the tests that are used must be accurate indicators of what students can do. In this, the need for performance measures is clear.

Dale Carlson, California State Department of Education

Is testing helping or hurting schools that are in trouble?

- Educators are doing a good job of testing only if they believe that students should be learning by rote or completing multiple-choice tests. If thinking were a skill, it could be taught and learned through rote methods and tested using current methods.

- Students learn through their senses, so testing should mirror the process of learning as well as the content that is learned; other testing methods should supplant multiple-choice tests. Students can demonstrate competence in an old-fashioned, sensible way: through oral assessment, which can catch the way students learn.

- Salient issues in testing today include: internal versus external control of tests, test reliability, use of portfolios, and testing on demand.

Pascal Forgione, Connecticut State Department of Education

Connecticut currently is introducing into its high schools a new generation of math and science tests, an action that is a response to the realization that current practices are inadequate for education systems that are planning for the 21st Century. The time has come to engage in a longitudinal testing program where much more time is spent assessing each student.

- Performance testing must integrate three often isolated elements—a common core of learning, preparation for life, and a global perspective—and draw from the fields of research, instruction, and assessment.

- A successful assessment program must contain tests that students consider essential, authentic, rich, engaging, active, and feasible.

- The basic model in Connecticut's program is that students will think while performing tasks. The observable tasks will indicate the level of thinking in which students are engaged.

ACCOUNTABILITY AND AT-RISK STUDENTS

The architects of many of the first reform policies did not consider the impact that their initiatives would have on students considered to be at risk. Although these policies usually were launched with the intent of raising the quality of education for all students, considerable evidence shows that these policies often further discourage at-risk students and act to push them out of the school system.

Recent attention to long-term demographic trends and their future impact on the nation's workforce has generated fresh concern about the consequences of reform. At the same time, members of the education community have questioned the use of

commercial tests to assess the achievement of at-risk students. They are concerned that testing bias may work against at-risk students, that a narrow focus on highly specific knowledge and basic skills may lead to instructional practices that are particularly harmful for the progress of these students, and that new methods of assessment also may have repercussions.

Panelists were asked to comment on the possible consequences connected to existing accountability systems for at-risk students and the longer term consequences connected to the implementation of expanded assessment systems and reporting practices.

Jeannie Oakes, University of California, Los Angeles

Educators and policy makers must create indicator systems that can be used to improve the educational situation of students who are historically at risk.

- Educational data on these students must be reported in such a way that their achievement is represented fairly. However, controlling and adjusting for background factors raise a variety of logistical and technical problems.

- Background variables must not be used to institutionalize lower expectations for certain schools.

Ruben Carriedo, San Diego Unified School District

Test scores from norm-referenced tests are used to evaluate many aspects of education in the San Diego schools.

- Although many of the district's parents, board members, and educators think that there currently is too much testing, other constituents are reluctant to accept new ways of assessing student performance. In particular, the minority community is suspicious of new measurement approaches, thinking that new measures may sidestep the issue of how their children are really doing.

- Constituents need to be educated about alternatives to norm-referenced testing.

- San Diego plans to replace CTBS with a shorter, norm-referenced test developed by the district. Portfolio assessment is planned for senior high social studies and middle school English and math.

Harriet Doss Willis, Southwest Regional Laboratory

In what sense can the term "minority" be considered a useful designation? Lumping students into one group can generate problems.

- Assessment tends to treat all students the same, regardless of their background or the actual instruction they receive. Students may be identified as underachievers as a result of monolingual instruction or individual learning styles.

- Aspects of a student's background—race, socio-economic status, family background—that might be used to adjust achievement results should not be made into excuses for failure to teach or learn, but should be used to develop new strategies for instruction and assessment.

Jerome Jones, St. Louis Public Schools

In the St. Louis school system, the issue of accountability does not center on testing, but on expectations: Teachers are encouraged to believe that all students can learn and that all teachers can teach.

- The St. Louis school system stresses traditional achievement measures in the belief that they create an equitable standard against which performance can be gauged.

- Continuing problems arise from an absence of commitment, a curriculum that was developed without reference to socio-economic conditions, and social agencies that are isolated from the education system.

Robert Rueda, University of Southern California

Current research at USC on literacy implies some of the pitfalls of using top-down indicators to address the problems of at-risk students.

- Outcome measures don't reflect the process of learning or how students interpret tasks. More local indicators should be used.

- Indicators must be tied to a theory of teaching, and pupil progress must be monitored. If indicators do not measure the important aspects of literacy, underestimation of students' abilities can result; if indicators measure the wrong aspects, overestimation can result.

- A student who is allowed to write in his native language or on authentic topics may show more ability or more potential than is indicated through top-down assessment.

James Catterall, CRESST, University of California, Los Angeles

Projects underway at CRESST are attempting to show the effects of recent reform policies on students who are considered at risk.

- A study that is examining tests required for graduation in four states indicates that the treatment of student failures varies widely, ranging from the humiliation of students to placing them in remedial programs.

- In the School Reform Assessment project, CRESST researchers are using interviews, student surveys, and examination of student transcripts to look at the consequences of increased course-taking requirements, particularly for students who are at the lower end of the achievement spectrum.

INTERNATIONAL EDUCATIONAL INDICATORS: THEIR INTERNATIONAL AND NATIONAL ROLES

Public concern for the quality of American education has been heightened by results from cross-national studies of educational achievement. The same type of concern is evident in other major industrial nations, where educators and policy makers are feeling pressure to improve their education systems. In addition, the pending formation of the European Community is fostering interest in enhancing the quality and comparability of international

educational systems.

To address these concerns and interests, the Organization for Economic and Cultural Development (OECD) is completing a project on the feasibility of developing an international educational indicator system and is proposing a two-year conceptualization and development phase. NCES is one of the American agencies that is interested in the development of such a system. OECD's effort is touching on many of the issues that arise in discussions about the development of state, district, and school-level indicator and reporting systems in the U.S.

Panelists were asked to examine current and anticipated development and use of international educational indicators.

Norberto Bottani, OECD

In May of 1988, OECD decided to implement an international program to identify educational indicators that will aid the evaluation of the quality of educational performance. Twenty-four countries are participating in the program.

- The majority of developing countries invest at least 5% of their GNP in education, yet do not have a systematic way of describing and evaluating their educational system.
- During the exploratory phase of the OECD program, each of five networks, or working groups, investigated an aspect of the topic: student flows, student achievement, school functioning, school facilities and resources, and the attitudes and expectations of teachers and students. Participating countries could be involved in any or all of the five networks.
- Conceptual problems have hindered the identification of indicators. Many European countries make little use of objective testing, and, thus, results from their exams may not be compatible with U.S. scores, which are based almost exclusively on multiple-choice or other objective test formats.
- At present it is impossible to produce a single indicator that applies to all 24 countries or that is acceptable to all participants.

Jeanne Griffith, National Center for Education Statistics

The idea of an international indicators project was first advanced by the U.S. An international system such as that being developed by OECD can tell us how the U.S. is doing in light of what is going on around the world.

- Three technical issues bedevil the development of indicator systems: Development of standards for comparison; how the information should be reported (data would be sufficiently detailed to inform decision makers, but not so overwhelming that it will defy simple reporting of findings); and how higher order thinking skills and subtopics should be measured.
- Issues of accountability should not be allowed to drive the composition of indicators. Indicators that are more difficult to measure and report should not be put aside in favor of indicators that can be measured easily.
- NCES has initiated research on educational indicators in the U.S. The Center's annual report on indicators, Condition of Education, is similar to the system being proposed by OECD, although the point of reference changes from time to time. In contrast, OECD will have a stable set of indicators over time.

Dean Jamison, University of California, Los Angeles

The World Bank has commissioned the development of indicators that relate to the health of sub-Saharan Africans. It is somewhat difficult for education to replicate the quantitative nature of many health indicators, but the system developed by the World Bank has many features that can be generalized to all indicators projects.

- In any indicator system it is necessary to "organize" the personnel involved. Different people will be interested in different types of indicators: Technicians and psychometricians favor well-proven, relatively easily provided indicators, while political leaders want different information. Whatever data are produced will be used regardless of the desires of politicians.
- At present, U.S. economic productivity has not been adversely affected by poor educational performance, but this is predicted to change over the next few decades. Some researchers have proposed that economic indicators be used to aggregate the potential cost of the U.S.'s educational inferiority to other nations.
- Few studies have linked educational indicators to economic output in OECD countries; such studies are needed in order to quantify the effect of education on economic activity.

William Schmidt, Michigan State University

The results of cross-national achievement should be placed in the context of national educational environments, or the results will be useless for providing sound policy direction.

- The alignment of data to requests for information often is incompatible, or, at best, an imperfect match. Any useful results must be content-specific and must be interpreted in light of concomitant covariates. The inclusion of Opportunity to Learn (OTL) indicators is crucial if the results of cross-national studies are to be used for policy reform.
- Recent IEA studies of math and science offer some clear findings in regard to U.S. education. These findings indicate that achievement in these areas is in proportion to the stress placed on studying the topics.
- In the IEA study of math, researchers found that in most countries the eighth-grade math curriculum is devoted mainly to algebra. Seventy percent of content in Japanese eight-grade math classes is focused on algebra, compared to 7% in U.S. math classes.
- In the IEA study of science, U.S. achievement levels were the lowest of all countries assessed. Sixty-eight percent of U.S. schools did not have a ninth-grade science requirement.

Summary of Questionnaire Responses

Damian Murchan
CRESST Fellow, Cornell University

Working group sessions, held on the afternoon of October 13, were designed to generate input from conference participants. These sessions focused on the issues that structured the conference agenda and that arose during conference proceedings.

To encourage and facilitate participant contributions to the working group sessions, participants were asked to complete a questionnaire prior to the sessions. Questionnaire items dealt with the development and use of educational quality indicators. Participants were asked to give specific consideration to implications for policy and policy makers, for practice and technical assistance, and for research and development. Session leaders used participants' responses to guide group discussion during the working group sessions.

Respondents included state agency heads, directors of state evaluation and assessment programs, local program evaluators, school board members, university faculty, education analysts, and research and development personnel.

The following summary systematically examines the questions posed and provides an overview of the written responses.

Potential Benefits of Indicator Systems

The first question asked participants to identify benefits that could be gained from the development and use of educational quality indicators and to note the levels of the education system that would benefit.

Two themes dominated respondents' thinking in this regard: accountability and providing information to policy makers. Accountability was seen as a primary benefit of indicator systems. Many respondents wrote that the demand for accountability is increasing and that indicator systems could provide a way of satisfying the demand. Some saw accountability as relating not to students or teachers, but to the educational programs themselves, implying that the designers of the program are under scrutiny. One person thought that an indicator system should be used to direct the operation of the education system to achieve agreed-upon goals.

Many respondents noted that indicators could be used to inform stakeholders. Routine assessment of indicators could generate interest and concern in public education and the multiple expectations and outcomes of schooling. Indicators could also indicate trends over time, thus providing data to policy makers that would facilitate implementation of remedial action. A clearly understandable set of indicators would maximize the chance that the public would be willing to spend the additional monies necessary to remediate deficient areas of the education system.

Fewer respondents thought that an indicator system would have immediate impact on classroom practices. Rather, the function of indicators was seen as aiding decisions at higher levels, such as in state legislatures and state evaluation agencies. Particularly, respondents indicated that indicator systems could be used to identify differential educational opportunity, which might lead to the improvement of educational outcomes for all students, especially the less fortunate. However, several respondents thought that indicator systems would produce more political benefits than benefits at the school or class level. One respondent noted that tangible gains could be made at the classroom level if performance measures were included in indicator systems.

Potential Dangers of Indicator Systems

The second question asked respondents to list potential dangers associated with the development and use of educational quality indicators.

The overriding concern among respondents was that the indicator system would be too narrow in scope and that the indicators included in the system would become the sole focus of educational efforts. Specifically, respondents alluded to the danger of an overemphasis on cognitive outcome measures, thereby leading to what one person called a "perversion of the educational purpose." Respondents felt that tried and tested measures would dominate the indicator system at the expense of equally important but more elusive process measures. Even within the confines of measuring cognitive outcomes, some people thought that it would be imperative to include performance measures if the system were to be credible.

Another concern centered on the danger that the data generated by indicators would be misinterpreted by politicians, the media, and parents. Many of the respondents warned that the data might be misused, precipitating crises in the education system: Teachers' professional authority could be undermined; students and teachers could be stigmatized and punished.

Some respondents felt that the very existence of the indicators might serve functions opposite to those intended, regardless of whether the data were used correctly. Conceivably, an indicator system could create a scenario where the losers would lose all. Any system of indicators should, according to several of the respondents, take into account the context of the school. Showing the education system to be lacking in terms of input and output might precipitate either of two general responses from the public: a movement to remedy deficiencies, or a lessening of support for educational improvement.

Methodological concerns also arose. Some dealt with test format: Responses suggested that no single-item format is sufficient and that if a mixed format is used, measures must necessarily be more complex. Other concerns were related to test administration: If site personnel are required to implement parts of the assessment, how will insufficient training be countered? One respondent was bothered by what he termed "normative reference group bias"; another questioned the probability of obtaining sufficient sample sizes to give the necessary level of detail. Many of these methodological concerns could become quite serious if measures are not continually updated and evaluated to ensure that they remain appropriate over time.

Improving the Quality of Education

The third question asked respondents to identify indicators that should be given priority in assessing and improving the quality of education.

The many indicators mentioned can be divided into categories relating to cognitive and social development, school and class environment, resource allocation and input, instructional practice,

and leading indicators used to predict outcomes. Respondents gave particular emphasis to assessing literacy, math, and science. Slightly less emphasis was given to measures of students' higher order reasoning and measures such as performance tests. Respondents indicated that the indicators linked to cognitive measures should give some indication of the requirements connected to the indicators; for example, if course grades are reported, the expected performance level in each course should be provided as well. In addition, indicators of excellence should be reported routinely for units such as schools and states; examples might be AP scores or performance on other advanced tests. Such scores could provide national bench marks that local systems could use to evaluate their own education systems.

Students' social development emerged as another area that merits attention. Respondents felt that indicators should measure maturity, adult functioning, and ability to cooperate in groups to achieve desired ends. Other similar measures mentioned by respondents were motivation level of the class, safety of the school environment, attendance, dropout rates, proportion of students graduating from high school, and proportion of students seeking further education.

Resources were seen to play an important role in determining school effectiveness. Respondents stressed that costs allocated to education, pupil-teacher ratio, and teacher workload and expectations should be among the factors assessed, since outcome measures alone cannot adequately describe functioning educational systems.

Measures of process variables and other related variables were also deemed useful for inclusion, according to respondents. One respondent pointed out that measures of opportunity to learn are vital for a proper assessment of cognitive outcomes. In general, respondents indicated that data should be sought that will help improve instruction; one example cited was the instructional time spent on specific tasks. Other factors that are known to correlate with achievement, such as the organizational structure of school and classroom, might also prove valuable in evaluating outcome scores and concomitant variables.

A few people indicated the desirability of including leading indices—variables that would enable the prediction of future educational quality—in indicator systems. Such indicators could pinpoint instructional practices or environmental variables that are associated with eventual dropout or identify course enrollment patterns associated with desired learning outcomes. Allied to this was a perceived need to look at students' performance after finishing or leaving school (as indicated by employment records), the proportion of students proceeding to post-secondary education, and the attitudes and expectations of teachers and students. The incorporation of such indicators might well necessitate the gathering of qualitative, ethnographic data.

Priorities at National, State, and Local Levels

Question four asked participants to determine differences in priorities at national, state, and local levels.

Respondents perceived a need at the national level for gross indicators of educational functioning that would serve two goals: aid in formulating policy and facilitating accountability. Policy indicators could provide data that would aid the optimum appropriation of funding to different structures, such as teacher training, if a deficit were unearthed in this area. Indicators could also help policy makers appraise the broad curricula being implemented in schools by surveying central tendencies and ranges of performance by geographic region. In addition, indicators gathered at a cross-national level could be useful in formulating curricular or other innovations in light of normative information from other systems. Respondents felt that international comparisons would be useful in measuring the performance of all students, a desirable function since these students will be crucial in providing future scientific and technological leadership and their abilities will have obvious implications for economic competitiveness.

One priority identified at the state level was the use of indicator systems to compare states and districts within states. State education personnel interested in comparisons with other states would find the data useful for policy formulation if the indicators were constructed in enough detail to suggest solutions for the deficiencies that indicators would identify, whether in terms of input or student performance. Accountability also was mentioned as a priority at the state level.

Many respondents thought that highly specific indicators would be needed at the local level as well; such indicators could facilitate on-the-spot remedial action by teachers. These respondents wanted to see the development of "improvement-oriented" indicators that would simplify the task of figuring out what to do with the information generated by the indicator system. Indicators of local teacher and administrator performance would fall into this category.

Four of the respondents reported that priorities should be the same at every level, pointing out the need at all levels for core skills. They indicated that the goals of education seem to be similar at all levels; thus, because many indicators depend on goals, no differences in priorities should occur.

Indicators with a Strong Foundation

The fifth question asked respondents to list indicators that have a strong foundation and that generate reliable data and reports.

The indicator cited most frequently was standardized achievement tests, particularly in the basic content areas, though the caveat was added that educators tend to fail in ensuring that the scores are interpreted correctly. Other responses regarded: school data on enrollment, dropout, retention, absenteeism, and staffing; indicators that track expenditure on education; aptitude data as measured by the SAT and ACT; indicators of opportunity to learn; and local descriptive indicators. Some respondents reported that no indicators presently have a sound foundation.

One person indicated that the education community should be wary in congratulating themselves on what seem to be valid and accurate indicators and gave as an example per-pupil expenditures: A dollar figure means different things depending on the context of the school. The same amount of money will not go as far in a cold climate as in a moderate climate because heating and maintenance costs will probably be higher where temperatures are colder. Thus, the per-pupil expenditure, which is usually calculated by dividing the school budget by the number of students served, fails to tell the full story.

Indicators That Need Better Methodology

The sixth question asked respondents to list indicators that need better methodology for collection, reporting, and use.

Most people thought that indicators relating to pupil achievement need critical attention. Those mentioned were: performance assessment in mathematics, science, and social studies; measures of higher order thinking; portfolio assessment for language arts; appropriate measures of fine art and music; and a

better measure of overall grade point average than is presently offered in survey research. In addition, one person thought that it would be useful to concentrate on developing better indicators of the number of advanced courses taken by students.

Measures of student persistence and participation in school were also deemed to require sound methodology. Some respondents perceived a need to clarify exactly what characterizes a dropout: Should information on the number of dropouts who then complete GED requirements be included? Similar indicators that were listed related to student attitudes, social and personal development, and career orientation.

Some process variables such as teacher performance, the performance of administrators, and the performance of units at the state and local levels were mentioned, as was measuring teacher professionalism.

Indicators That Need To Be Rethought

Respondents were asked next to identify indicators that need reconceptualization.

These, in the order of the frequency with which they were mentioned, were: student performance in basic subjects, social studies, the fine arts, and philosophy; problem-solving skills and applying concepts; leading indicators that would help predict future achievement and identify at-risk students; and drop-out tabulation. Also cited were teaching strategies and other process variables, motivation, discipline, time spent on learning-related activities outside school, and the "wall chart."

Barriers to Responsible Use

Question eight related to barriers to the appropriate and responsible use of indicators.

Respondents saw technical and methodological constraints as the dominant threat to the proper use of indicators. Respondents identified problems such as a deficient research base, inadequate conceptualization and consequent ambiguity of the measures, failure to see that indicators need to serve different functions at different levels, unreliable measurement, and insufficient attention paid to interpretation of results. Respondents indicated that many of these problems might be alleviated if sufficient resources were expended in designing the indicator system, but several respondents suspected a general unwillingness to spend the money necessary to resolve the technical problems and mentioned that state legislators and other politicians extoll the education community to use less costly (and possibly less valid) indicators. This perceived pressure to come up with quick solutions was a source of concern to six of the respondents.

Resistance from teachers was thought to be a potential barrier by some respondents, particularly resistance connected to the possibility that the use of indicators would become a high-stakes venture for schools whose funding or student enrollment could be adversely affected by negative findings. Several of the respondents noted the difficulty of obtaining consensus as to what the indicators should be. Respondents also pointed out that many people believe that test scores are comprehensive and accurate indicators of educational effectiveness; convincing them to look beyond a limited set of cognitive outcome scores to a set of indicators that incorporate input and process-oriented factors could be a difficult task. One respondent wrote, "As long as people, especially educationalists, believe that aptitude is a static quality and that learning is a piecemeal, fragmented, sequential process, we'll continue to count the equivalent of bumps on students' skulls." Even for educational practitioners, the job won't necessarily end when "scores" are determined for any set of indicators that has been developed. Some, if not many, of these indicators will not be accompanied by solutions or instructions for improvement. Thus, the experience and adaptability of users to make use of findings may be an important element in the ultimate worth of an indicator system.

Implications for the Education Community

The final question asked participants to determine the implications of educational quality indicators for (a) policy and policy makers, (b) practice and technical assistance, and (c) research and development.

The overriding issue in regard to policy and policy makers was that policy makers should be cognizant of the potential consequences, positive and negative, of implementing any indicator system. Respondents perceived a need to realize the power of the information yielded by the indicators. Several respondents stressed that this information, when interpreted correctly, could lead to improvements in the way children are educated; misunderstood, it could be highly detrimental for the system as a whole. Particular attention should be paid to the consequences of an accountability use of indicators for low-achieving and at-risk students. One respondent held that overemphasis on the findings, with consequent stress on the education system, might actually invalidate their use altogether.

On a more positive note, a few respondents thought that the information yielded could prove useful in appraising teacher education and bringing about an increased emphasis on training teachers to develop students' thinking skills. Indicators could also provide additional data for the formulation of appropriate standards and expectations.

Several respondents noted that the dissemination of information was crucial to ensure the success of an indicator system. In addition, the need to implement a process to develop follow-up procedures was mentioned. However, if schools are forced to act on findings, initial teacher support for indicators could turn into indifference, if not outright opposition. On a less ominous note, another respondent wrote in favor of indicator systems, mentioning their value for school restructuring and the development of better curricula. When used at the national level, a comprehensive set of valid indicators would necessitate a dramatic broadening in the definition of national assessment, which would demonstrate clearly to policy makers that no one test will answer all the relevant questions and that there is no simple answer to the questions they pose.

In regard to practice and technical assistance, respondents identified two main issues. First and foremost, respondents felt that the designers of indicators must explain the results. It would not be acceptable to leave their interpretation in the hands of policy makers, chief state school officers, school superintendents, or teachers. Rather, educational improvement can be attained only through careful explanation of the data. Some respondents maintained that the resulting data bases should be easily accessible so that the research community may examine the data freely, which would ensure that the interpretation given to practitioners would stand up to professional scrutiny. Although ultimately it is up to individuals at the classroom level to implement change, the direction and support for change should be embedded in the indicator system itself by means of careful explanation of findings.

Providing such an array of indicators to education practitioners must be accompanied by the implementation of more appropriate content in teacher education courses and the provision of in-service courses for practicing teachers. These courses could prepare the practitioners for work in an environment where data would be routinely gathered; these courses would train teachers to utilize the data to improve their teaching strategies.

An indicator system, according to the respondents, would only be successful if it could provide feedback on effective practices to teachers and policy makers. Indicators could illuminate model systems and explain the relationships between certain input and process variables and subsequent outcomes. Over time, "before and after" analyses of school systems could be developed, whereby the interpretation of data would be provided in relation to a system before and after modifications were made on the basis of initial data. Such models would complement the policy of explaining and interpreting results.

In terms of research and development, respondents reported that the whole process of developing indicators needs considerable attention. Many recommended that the development of indicator systems should involve more school personnel from the outset, so that the systems will reflect the realities of the school and classroom situation—such a system might encourage teachers to collect their own data. In addition, respondents perceived a need to ascertain what parents and business want in terms of measuring a quality education. Specifying and quantifying information from varied sources would do much to ensure the adoption and ultimate success of indicator systems.

Many of the issues raised during the conference suggested that students' learning styles should be considered in the design of indicators. Thus, a foray should be made into cognitive and educational psychology in order to increase the research base on which the indicators will ultimately rest. Most respondents stated that broader indicators of student performance should be developed. These indicators should include not only performance measures, but measures that highlight student attitude, motivation, and interest. Though one respondent suggested exploration of a composite indicator that would do for education "what GNP does" for economic analyses, most envisaged an indicator system as a web-like structure that would be all encompassing but would not be expressible in some tidy coefficient such as an education version of GNP. Most respondents noted that each and every indicator should be constantly updated in response to ongoing research.

Several respondents pointed out that research and development efforts depend on open access to data. They argued that access is imperative if the research community is to ensure that the measures developed are psychometrically pure and relevant to the purposes for which they are designed. The whole issue of test or instrument development should be one that especially concerns the measurement community, a community who seem increasingly committed to solving the problems associated with moving away from traditional pencil-and-paper test formats.

Publications Available from the UCLA Center for the Study of Evaluation

Recent Technical Reports

Comparing Four Statistical Packages for Hierarchical Linear Regression: GENMOD, HLM, ML2 and VARCL
Ita G.G. Kreft, Jan de Leeuw and Kyung-Sung Kim
CSE Technical Report 311, 1990 ($4.00)

Report on Content Definition Process in Social Studies Testing
Ernest R. House and Nancy Lawrence
CSE Technical Report 310, 1990 ($3.00)

Patterns in Teacher Reports of Topic Coverage and Their Effects on Math Achievement: Comparisons Across Years
Bokhee Yoon, Leigh Burstein, Zheng Chen and Kyung-Sung Kim
CSE Technical Report 309, 1990 ($2.50)

Comparing State and District Test Results to National Norms: Interpretations of Scoring "Above the National Average"
Robert L. Linn, M. Elizabeth Graue and Nancy M. Sanders
CSE Technical Report 308, 1990 ($5.00)

"Inflated Test Score Gains": Is It Old Norms or Teaching the Test?
Lorrie A. Shepard
CSE Technical Report 307, 1990 ($2.50)

Duplex Design: Giving Students a Stake in Educational Assessment
R. Darrell Bock and Michele F. Zimowski
CSE Technical Report 306, 1990 ($2.50)

Analyses of Procedures for Assessing Content Coverage and Its Effects on Instructional Assessment
Leigh Burstein, Zheng Chen and Kyung-Sung Kim
CSE Technical Report 305, 1989 ($4.50)

R&D Priorities for Educational Testing and Evaluation: The Testimony of the CRESST National Faculty
Joan L. Herman (Editor)
CSE Technical Report 304, 1989 ($3.00)

Using Multilevel Analysis to Assess School Effectiveness: A Study of Dutch Secondary Schools
Ita G.G. Kreft
CSE Technical Report 303, 1989 ($2.50)

Has Item Response Theory Increased the Validity of Achievement Test Scores?
Robert L. Linn
CSE Technical Report 302, 1989 ($3.00)

Developing Indicators of Student Coursework
Lorraine M. McDonnell and Tor Ormseth
CSE Technical Report 301, 1989 ($3.00)

The ACOT Report Card: Effects on Complex Performance and Attitude
Eva L. Baker, Joan L. Herman and Maryl Gearhart
CSE Technical Report 300, 1989 ($1.50)

Technology Assessment: Policy and Methodological Issues
Eva L. Baker
CSE Technical Report 299, 1989 ($2.50)

Reporting for Effective Decisionmaking
Joan L. Herman, Lynn Winters and Shari Golan
CSE Technical Report 298, 1989 ($2.50)

Model-Based Ranking of Schools
Ita G.G. Kreft and Jan de Leeuw
CSE Technical Report 297, 1989 ($2.50)

Publications Available from the UCLA Center for the Study of Evaluation

School Dropouts: Here Today, Here Tomorrow
James S. Catterall
CSE Technical Report 296, 1989 ($2.50)

Higher Order Assessment and Indicators of Learning
Eva L. Baker
CSE Technical Report 295, 1989 ($2.00)

A Classification of Sentences Used in Natural Language Processing in the Military Services
Merlin C. Wittrock
CSE Technical Report 294, 1989 ($2.50)

Survey on ECTA Chapter I Evaluation Regulations
Sharon Johnson-Lewis (Editor)
CSE Technical Report 293, 1989 ($4.00)

Instructional Sensitivity in Mathematics Achievement Test Items: Application of a New IRT-Based Detection Technique
Bengt O. Muthen, Chih-Fen Kao and Leigh Burstein
CSE Technical Report 292, 1988 ($3.00)

Cultural Literacy and Testing
Ernest R. House, Carol Emmer and Nancy Lawrence
CSE Technical Report 291, 1988 ($5.00)

Can We Fairly Measure the Quality of Education?
Eva L. Baker
CSE Technical Report 290, 1988 ($1.50)

Increasing the Utility of Information Systems in Schools: Lessons from the Literature
Joan L. Herman and Shari Golan
CSE Technical Report 289, 1988 ($5.00)

Directly Comparing Computer and Human Performance in Language Understanding and Visual Reasoning
Eva L. Baker, Elaine L. Lindheim and Josef Skrzypek
CSE Technical Report 288, 1988 ($2.00)

A Contrast Between Computer and Human Language Understanding
Eva L. Baker and Elaine L. Lindheim
CSE Technical Report 287, 1989 ($2.00)

Instructionally Sensitive Psychometrics: Applications to the Second International Mathematics Study
Bengt O. Muthen
CSE Technical Report 286, 1988 ($4.50)

Implementing STAR: Sensible Technology Assessment/Research
Eva L. Baker and Joan L. Herman
CSE Technical Report 285, 1988 ($1.50)

The Role of Symbolic Representation in Achievement and Instruction
Noreen Webb, Sen Qi and John Novak
CSE Technical Report 284, 1989 ($5.50)

Mandated Tests: Reform or Quality Indicator?
Eva L. Baker
CSE Technical Report 283, 1988 ($2.00)

Dimensions of Thinking: Implications for Testing
Robert L. Linn
CSE Technical Report 282, 1988 ($3.50)

Multiple Choice Questions as a Diagnostic Tool
Pinchas Tamir
CSE Technical Report 281, 1988 ($2.50)

Conversations on Evaluation Utilization
Marvin Alkin (Editor)
CSE Technical Report 280, 1988 ($5.50)

Designating Winners: Using Evaluation in School Recognition Programs
Edward Wynne (Editor)
CSE Technical Report 279, 1988 ($5.50)

Standards and School Dropouts: A National Study of the Minimum Competency Test
James S. Catterall
CSE Technical Report 278, 1987 ($3.00)

The Texas Teacher Test
Lorrie A. Shepard and Amelia E. Kreitzer
CSE Technical Report 277, 1987 ($1.00)

A Case Study of the Texas Teacher Test
Lorrie A. Shepard, Amelia E. Kreitzer and M. Elizabeth Graue
CSE Technical Report 276, 1987 ($7.50)

State-by-State Comparison of Student Achievement: The Definition of the Content Domain for Assessment
Robert L. Linn
CSE Technical Report 275, 1987 ($1.50)

Evaluation for School Improvement: Try-out of a Comprehensive School-Based Model
Joan L. Herman
CSE Technical Report 274, 1987 ($1.50)

Definition of Content in Social Studies Testing: Conceptual Content Assessment Report
Ernest R. House, Carol Emmer, Elaine Kolitch, Barbara Waitz and Eva L. Baker
CSE Technical Report 273, 1987 ($5.00)

Issues in Intelligent Computer-Assisted Instruction: Evaluation and Measurement
Harold F. O'Neil, Jr. and Eva L. Baker
CSE Technical Report 272, 1987 ($2.50)

Using Item Specific Instructional Information in Achievement Modeling
Bengt O. Muthen
CSE Technical Report 271, 1987 ($2.50)

Translation Among Symbolic Representation in Problem Solving
Richard J. Shavelson, Noreen M. Webb, Michal Shemesh and Jin-Wen Yang
CSE Technical Report 270, 1987 ($2.00)

The Role of Symbol Systems in Problem Solving: A Literature Review
Richard J. Shavelson, Noreen M. Webb, Michal Shemesh and Jin-Wen Yang
CSE Technical Report 269, 1986 ($2.00)

Some Uses of Structural Equation Modeling in Validity Studies: Extending IRT to External Variables Using SIMS Results
Bengt O. Muthen
CSE Technical Report 268, 1986 ($2.50)

Speed and Accuracy of Word Decoding and Recognition
Robert L. Linn, S.W. Valencia and K.E. Ryan
CSE Technical Report 267, 1987 ($3.50)

Reading Assessment: Practice and Theoretical Perspectives
Robert L. Linn and S.W. Valencia
CSE Technical Report 266, 1986 ($2.50)

Educational Quality Indicators in the United States: Latest Developments
Leigh Burstein
CSE Technical Report 265, 1986 ($1.50)

Monograph Series

Assessing Student Achievement: A Profile of Classroom Practices
D.W. Dorr-Bremme and Joan L. Herman
CSE Monograph 11, 1986 ($11.00)

Evaluation in School Districts: Organizational Perspectives
Adrianne Bank and R.C. Williams (Eds.)
CSE Monograph 10, 1981 ($7.50)

Values, Inquiry and Education
H.D. Gideonse, R. Koff and J.J. Schwab (Eds.)
CSE Monograph 9, 1980 ($11.00)

Toward a Methodology of Naturalistic Inquiry in Educational Evaluation
E. Guba
CSE Monograph 8, 1978 ($4.50)

The Logic of Evaluative Argument
Ernest R. House
CSE Monograph 7, 1977 ($4.50)

Achievement Test Items: Methods of Study
C. Harris, A. Pearlman and R. Wilcox
CSE Monograph 6, 1977 ($4.50)

Resource Papers

Improving Opportunities for Underachieving Minority Students: A Planning Guide for Community Action
Josie G. Bain and Joan L. Herman
CSE Resource Paper 8, 1989 ($11.00)

Designing and Evaluating Language Programs for African-American Dialect Speakers: Some Guidelines for Educators
Pauline E. Brooks
CSE Resource Paper 7, 1987 ($2.00)

A Practical Approach to Local Test Development
James Burry, Joan L. Herman and Eva L. Baker
CSE Resource Paper 6, 1984 ($3.50)

Analytic Scales for Assessing Students' Expository and Narrative Writing Skills
Edys S. Quellmalz and James Burry
CSE Resource Paper 5, 1982 ($3.00)

Criteria for Reviewing District Competency Tests
Joan L. Herman
CSE Resource Paper 4, 1982 ($2.00)

Issues in Achievement Testing
Eva L. Baker
CSE Resource Paper 3, 1982 ($2.50)

Evaluation and Documentation: Making Them Work Together
James Burry
CSE Resource Paper 2, 1982 ($2.50)

An Introduction to Assessment and Design in Bilingual Education
James Burry
CSE Resource Paper 1, 1982 ($3.00)

New from CRESST

The Undergraduates

C. Robert Pace

Most college students spend a great amount of time on their academic work and feel they have made substantial progress toward important goals. In fact, most students report that they are highly satisfied with their undergraduate experiences.

These conclusions are among those presented in *The Undergraduates*, a publication that offers information often ignored in the assessment of higher education: the students' perspective of their undergraduate education. The book focuses on the scope and quality of effort that students invest in their undergraduate experiences and the amount of progress students think they make toward educational goals.

The results and conclusions presented in *The Undergraduates* are drawn from student responses to the College Student Experiences Questionnaire, an instrument that has been used to survey more than 25,000 undergraduate students in the past several years.

Results are reported for each of five types of institutions: research universities, other doctoral universities, comprehensive colleges and universities, traditional liberal arts colleges, and highly selective liberal arts colleges.

Copies of *The Undergraduates* are $19.50 each.

Improving Large-Scale Assessment

Pamela Aschbacher and Eva L. Baker, Editors

Improving Large-Scale Assessment presents a series of reports developed by CSE/CRESST to provide state and local educational testing officers with guidelines for ensuring the technical quality of large-scale assessment programs. *Improving Large-Scale Assessment* is the product of a unique task force brought together by CRESST to identify issues and needs and to provide options for improving testing and evaluation practices. Reports are housed in a three-ring binder.

The first installment of *Improving Large-Scale Assessment* is "Guidelines for the RFP Process." This report presents a systematic model for developing an assessment RFP. It contains discussions of basic issues, approaches to planning the RFP, communicating with bidders, the RFP structure, and the review process. "Guidelines" outlines the pros and cons of the test procurement process and shares the viewpoints and experiences of CSE/CRESST personnel.

"Guidelines" is shipped with the binder. Reports on additional topics will be issued over the next several years.

The first copy of *Improving Large-Scale Assessment* and "Guidelines for the RFP Process" is free of charge to school districts or state testing offices; additional copies are $10.00 each.

Publications by CRESST Staff

Testing and Cognition

Merlin C. Wittrock and Eva L. Baker, Editors

Testing and Cognition presents an up-to-date look at advances in cognitive psychology and their implications for the assessment of students. The chapters in *Testing and Cognition* cover a wide range of topics that relate to metacognition, motivation, and other affective processes, to particular subject matter assessment, and to implications for practice. The authors are acknowledged leaders in the field.

Testing and Cognition includes specific examples of the relationship of theory to practice in the subject matter areas of mathematics and history. These examples provide models that can be used by practitioners in a wide range of fields.

Available in 1990 from Prentice-Hall, Englewood Cliffs, New Jersey.

Making Schools Work for Underachieving Minority Students

Josie G. Bain and Joan L. Herman, Editors

The academic performance of disadvantaged students is of increasing concern to the education community and the public. *Making Schools Work for Underachieving Minority Students* explores the problems that these students face and offers suggestions intended to better their educational opportunities and increase their academic achievement.

The contributors to *Making Schools Work for Underachieving Minority Students* are distinguished researchers, practitioners, and policy makers who are committed to improving education for at-risk students. They represent a range of viewpoints and experience and provide a comprehensive assessment of the current status of education for these students. *Making Schools Work* grew from the proceedings of a national conference sponsored by CSE/CRESST.

Available in 1990 from Greenwood Press, Westport, Connecticut.

Multilevel Analysis of Educational Data

R. Darrell Bock, Editor

Researchers have long been aware of the need for improved analysis methods for certain aspects of educational research, including surveys of curricular goals, examination of the effects of large-scale testing programs, and, particularly, the evaluation of school effectiveness. Such studies require hierarchical sampling designs that can accommodate data that is collected at multiple levels of the education system. Information from these studies can better reflect the relationships among ability and performance, teaching and learning, and policy and practice.

Multilevel Analysis of Educational Data provides an excellent introduction to the field and a guide to related literature. Contributers discuss methodology, application, and analysis of multilevel data. The papers in this collection were first presented at a research conference sponsored by CSE/CRESST and NORC.

Available from Academic Press, San Diego, California. Published in 1989.

Program Evaluation Kit, Second Edition

Joan L. Herman, Series Editor

The *Program Evaluation Kit* is a practical guide to planning and conducting a program evaluation. The step-by-step format includes tips, exercises, measurement instruments, and data collection forms. The *Kit* covers every technique necessary to evaluate any program and answers hundreds of questions that evaluators in all fields ask about research design, statistics, and performance measurement.

The nine volumes are written in non-technical language and feature examples from the fields of education, management, health, and social services—making the *Program Evaluation Kit* a valuable resource for a broad range of professions.

Available from Sage Publications, Inc., Newbury Park, California. Second Edition published in 1989.

Center for the Study of Evaluation
Eva L. Baker, Director
Joan L. Herman, Associate Director

Center for Research on Evaluation, Standards, and Student Testing
Eva L. Baker, Co-Director
Robert L. Linn, Co-Director

Evaluation Comment is published by the UCLA Center for the Study of Evaluation (CSE) and the Center for Research on Evaluation, Standards, and Student Testing (CRESST). CSE is a unique organization that is devoted to research, development, training, and dissemination in the field of testing and evaluation. Core support for CSE comes from the U.S. Department of Education's Office of Educational Research and Improvement (OERI), which funds CRESST. Substantial funding from a number of sources in the technology arena supports CSE's Center for Technology Assessment (C/TA). In addition, CSE provides service under contract to a wide variety of clients in the public and private sector.

Each issue of *Evaluation Comment* presents discussion on topics that deal with issues of theory, procedure methodology, and practice. CSE requests contributions for specific topics from recognized experts; unsolicited manuscripts are not accepted. Subscription to *Evaluation Comment* is free of charge; other publications are available for a nominal charge that covers the costs of printing. If you would like to receive *Evaluation Comment* and are not on our mailing list, please write to the CSE Dissemination Office, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, California, 90024-1522.

Production of *Evaluation Comment* is supported in part by a grant from the Office of Educational Research and Improvement, U.S. Department of Education. However, the opinions expressed in this publication do not necessarily reflect the position or policy of that agency, and no official endorsement by that agency should be inferred.

Order Form

Attach additional sheet if more room is needed

CSE Technical Reports

Report Number	Title	Number of copies	Price per copy	Total Price

CSE Monographs

Monograph Number	Title	Number of copies	Price per copy	Total Price

CSE Resource Papers

Paper Number	Title	Number of copies	Price per copy	Total Price

The Undergraduates Number of copies at $19.50 each _____ Total Price _____

Improving Large-Scale Assessment

First copy is free to school districts and state testing offices — additional copies are $10.00 each

Please send a free copy ☐ Number of copies at $10.00 each _____ Total Price _____

POSTAGE & HANDLING
(Special 4th Class Book Rate)

Subtotal of $0 to $10 add $1.50
of $10 to $20 add $2.50
of $20 to $50 add $3.50
over $50 add 10% of Subtotal

ORDER SUBTOTAL _____

POSTAGE & HANDLING (scale at left) _____

California residents add 6.5% _____

TOTAL _____

Your name & mailing address—please print or type:

Orders of less than $5.00 must be prepaid

☐ Payment enclosed ☐ Please bill me

MAIL TO:

CSE Dissemination Office
UCLA Graduate School of Education
405 Hilgard Avenue
Los Angeles, CA 90024-1522

EVALUATION COMMENT
UCLA CSE

Center for the Study of Evaluation
UCLA Graduate School of Education
405 Hilgard Avenue
Los Angeles, California 90024-1522

Non-Profit
U.S. Postage
PAID
UCLA